Model Archive Summary for Suspended-Sediment Concentration at Station 11455146

Liberty Cut at Little Holland Tract Near Courtland, California

This model archive summary summarizes the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC timeseries for the period of record 1/31/2014 to 1/23/2019. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.07/2016.10 (USGS, 2016) and USGS Techniques and Methods, book 3, chap C4 (Rasmussen and others, 2009). This summary and model archive is in accordance with Attachment A of Office of Water Quality USGS Technical Memorandum 2015.01 (USGS, 2014).

Site and Model Information

Site number: 11455146
Site name: Liberty cut at Little Holland Tract Nr Courtland, California
Location: Lat 38°19’43.86", long 121°40’03.11" referenced to North American Datum of 1983, CA, Hydrologic Unit 18020163
Equipment: A YSI EXO2 sonde began logging turbidity on January 31, 2014.

Model number: 11455146.SSC.WY14.1
Model calibration data period: October 22, 2015 to October 24, 2018 with the sample from January 23, 2019 not included in the dataset.
Model application date: January 31, 2014 to January 23, 2019.
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Physical Sampling Details and Sediment Data

All sediment data were collected using USGS protocols (USGS, 2006) and are stored in the National Water Information System (NWIS) database https://waterdata.usgs.gov/nwis. Discrete, boat-based sample collection for the sediment monitoring typically occurs in all seasons and between 6-12 times per year total while targeting storm events with high flow & sediment concentration during the winter and spring as well as sampling during summer low flow periods, spanning the range of turbidity conditions. Sample collection varied year to year while the sonde was deployed, with an average of 8 per year.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). Sediment samples represent the discharge-weighted concentrations of the stream cross section. The Equal Discharge Increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge weighted suspended-sediment samples were collected. Each sampling vertical is located at the centroid of increments representing 20% of the total flow (5 verticals). Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples to represent the average cross section because velocities are not always isokinetic (based on Table 4-5 from TWRI09A4, USGS 2006). A boat-based discharge measurement was collected immediately before sampling to determine the location of each vertical.

Technicians collected samples using either a FISP US D-74 or D-96 depth-integrated suspended-sediment sampler. The average channel depth is just under 15 feet and the thalweg is usually between 18-20 feet. Station velocities ranged from -0.9 to + 5.2 ft/sec. Sediment at this station is mostly fines (97% on average from sand/fine analysis) and potential sampling bias due to non-isokinetic sampling is considered minimal.

Samples were analyzed by the USGS sediment lab in Santa Cruz, California. All samples were analyzed for sediment concentration (mg/L) by the filtration method and many samples were also analyzed for the percentage of fines (<0.062 mm). Th sand/fine break analysis can be used to identify dataset variability and potential outliers and shows that this station is composed of mostly fines (97% on average). The depth integrated samples collected from each of the 5 verticals were not composited for analysis but were instead analyzed individually. This method of individual analysis is for quality control purposes because of rapidly changing, tidal conditions. Rapidly changing conditions in a tidal estuary could cause discrepancies in the sediment concentration between verticals so this was evaluated. Once the SSC from each vertical is validated, the set average SSC from the 5 depth integrated verticals in the cross section was computed to use in the calibration model. In rare occasions where the SSC at a vertical was deemed erroneous, a manual average was computed from fewer than 5 verticals and notes applied to the database. This occurred on 10/22/2015, 11/23/2015, 1/11/2016, 2/19/2016, 5/17/2016, 10/3/2017, and 6/12/2018 when the cross-sectional average was computed fewer than 5 verticals because outliers were identified. Evaluation of the dataset confirms that suspended-sediment concentrations laterally across the channel is mostly uniform and the outliers were due to sampling errors such as the sampler hitting the bed and biasing the sample. Outliers were identified when the SSC at a particular vertical exceeded 2 standard deviations from the average cross section.

All sediment data and associated metadata were reviewed and marked as approved in the USGS National Water Information System (NWIS) Water-Quality System database (QWDATA) before using in the calibration model. Publicly available field/lab sediment data can be found at: https://waterdata.usgs.gov/nwis/uv?site_no=11455146.

Surrogate Data

Continuous 15-minute turbidity data and discharge data were collected and computed by the USGS California Water Science Center and evaluated as possible explanatory variables for SSC. Turbidity data were measured using a YSI EXO2 sonde (and EXO turbidity sensor) and reported in Formazin Nephelometric Turbidity Units (FNU). Turbidity data began logging on January 13, 2014 up until the sonde was removed from the station on January 23, 2019 @ 1130. All surrogate turbidity data were computed, reviewed, and approved following methods in Wagner and others (2006). Discharge data were collected, computed, reviewed, and approved by the USGS California Water Science Center. Methods to compute discharge follow Levesque and

Oberg (2012). The 15-minute timeseries surrogate data were retrieved from NWIS-TS and are located at https://waterdata.usgs.gov/nwis/uv?site_no=11455146.

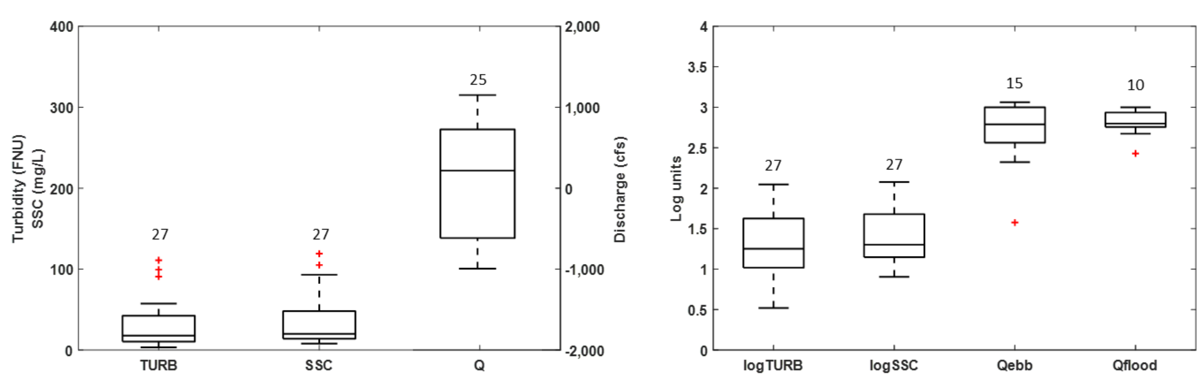
Model Calibration Dataset

The approved time-series data spanning the dates of the sediment constituent dataset were retrieved from NWIS-TS (Rasmussen and others 2009). The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate data with the discrete sediment data (Domanski and others 2015). Turbidity and discharge values were selected as a match for each sediment sample observation from a matching window of +/- 15 minutes for turbidity and discharge. The SAID manual is found at https://pubs.er.usgs.gov/publication/ofr20151177.

The sample from 9/21/2016 was flagged in SAID and was left out of the final dataset because it was not consistent with the previous day (with similar turbidity) and was collected during a changing tide (near slack). There was a data gap in the turbidity record during sample collection on 1/31/2018. Additionally, there were no surrogate values to pair with the sample collected on January 23, 2019 at 1343 during very high turbidity conditions; the final approved time-series ended prior to this sample. Thus, although there were 30 representative cross-sectional samples collected, only 27 were paired with surrogate turbidity data. Additionally, while there were turbidity data for the samples on 5/4/2017 and 1/9/2018, there were data gaps in the discharge record during those times.

Regression Model Development

Multiple models were evaluated including simple linear regression (SLR) and multiple linear regression (MLR). The most common estimation technique is SLR, but MLR is an alternate tool for computing SSCs when the SLR *MSPE* statistic is larger than 20 percent (Rasmussen and others, 2009). The calibration dataset is composed of 23 concurrent turbidity, SSC, and discharge measurements. Boxplots are shown below and note that due to negative tidal discharge values during the flood tide, ebb and flood values are shown separately with the absolute values shown during flood tides.



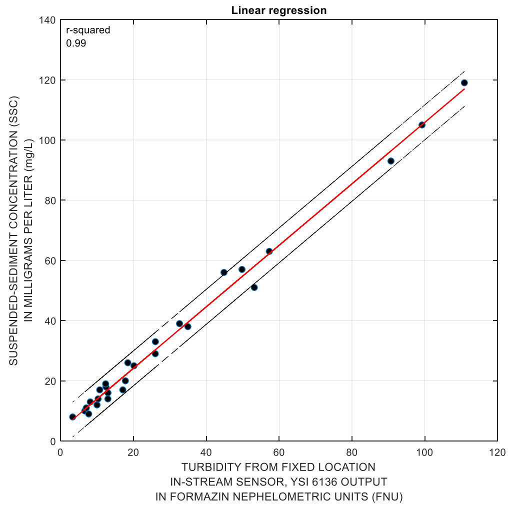
Model diagnostics and plots for model review were output using the combination of Matlab, SAID, and the R environment (R Core Team, 2018). The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the statistical diagnostics and guidance to evaluate them amongst the models. The best model was chosen based on R^2, significance tests (p-values), model standard error, correlation of explanatory variables, residual plots, variance inflation factor (VIF), and PRESS (prediction error sum of squares) statistics

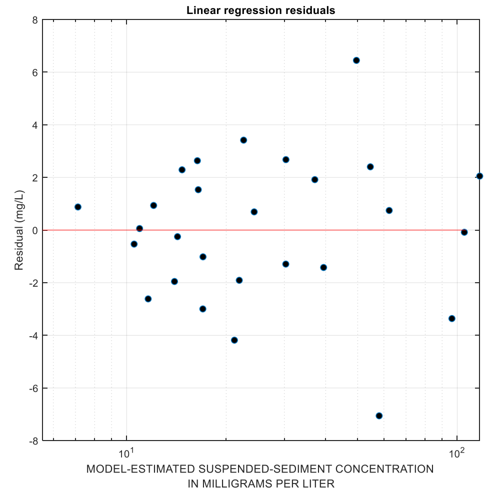
A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), and Model 5) $\log_{10}$ model with two explanatory variables (turbidity and absolute values of discharge). This allowed for inclusion of 25 discharge observations, ten of which were otherwise excluded with a multi-log regression due to negative discharge during flood tides. Diagnostic statistics are summarized below for the five models evaluated. The linear SLR model (model 1) is highlighted as the best option based on the highest R^2 and the lowest error statistics. Note that the RMSE for the log models are not comparable to the non-transformed models. Discharge was not considered further as a second surrogate (in addition to turbidity) because it was not significant to the model (p-value < 0.05) nor was it warranted by the MSPE statistic.

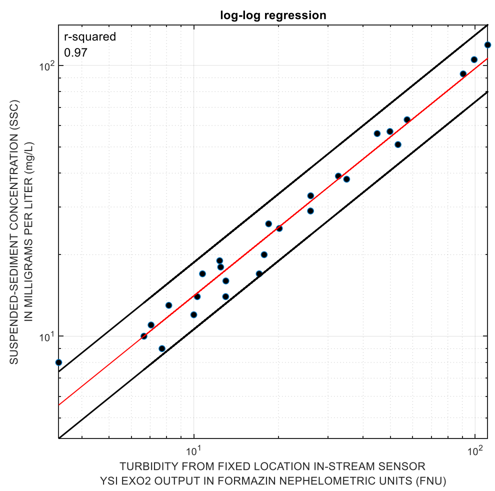
No.	R^2	R^2_a	RMSE	PRESS	MSPE	N	(type)
Model 1	0.99	0.99	2.8	230	8.2	27	linear
Model 2	0.97	0.97	0.1	0.1	13.9	27	log
Model 3	0.99	0.99	2.8	233	8.2	27	repeated median
Model 4	0.99	0.99	2.8	213	8.1	25	multi-linear
Model 5	0.97	0.97	0.06	26.6	14.6	25	multi-log ABS

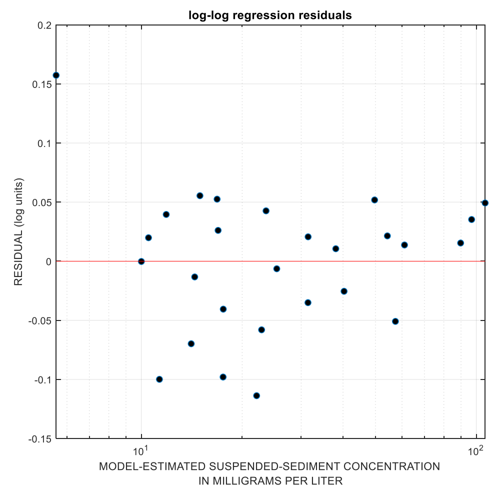
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the final model were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to – 3 range are considered potential outliers and were investigated. The studentized residuals were reviewed from the SAID output report and none of the samples collected were deemed outliers.

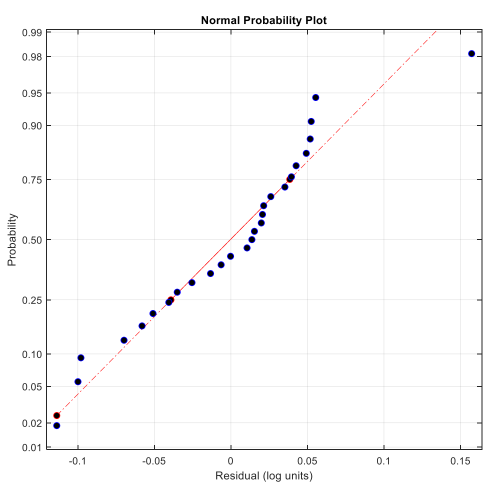
Of the SLR models, the linear regression model residual plots indicated a more homoscedastic pattern (constant variance) and a more normal distribution compared to the $\log_{10}$-transformed model. A comparison of the regression plots and residual plots are shown below for both model 1 and model 2.

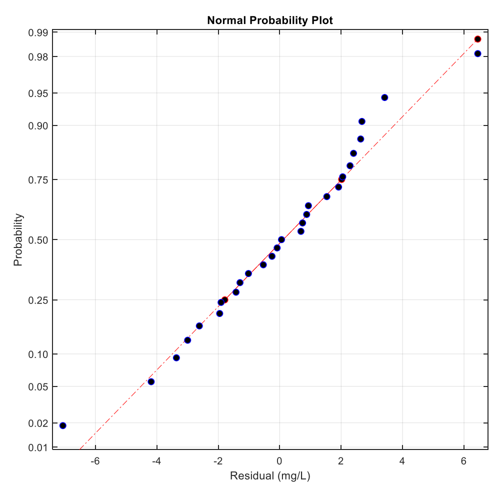

Linear regression
r-squared
0.99
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION
IN-STREAM SENSOR, YSI 6136 OUTPUT
IN FORMAZIN NEPHELOMETRIC UNITS (FNU)


Linear regression residuals
Residual (mg/L)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER


log-log regression
r-squared
0.97
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION IN-STREAM SENSOR
YSI EXO2 OUTPUT IN FORMAZIN NEPHELOMETRIC UNITS (FNU)


log-log regression residuals
RESIDUAL (log units)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER


Normal Probability Plot
Probability
Residual (log units)


Normal Probability Plot
Probability
Residual (mg/L)

Model Summary

The final simple linear regression model is based on 27 concurrent measurements of cross sectional suspended-sediment concentration samples and turbidity collected just over 3 water years from October 22, 2015 to January 23, 2019. USGS (2016) *recommends* a minimum of 36 paired observations, however the turbidity sensor was discontinued. Nonetheless, this model has very good agreement, low error, and model validation follows the guidelines in the Office of Water Quality Technical Memorandum 2016.10 for approval of surrogate regression models. The model information is shown below with basic model information, regression coefficients, correlation and summary statistics:

Linear Regression Model	**Coefficient of Determination (R^2)**
$SSC = 3.74 + 1.02 * Turb$	0.99

where

SSC is suspended-sediment concentration, in milligrams per liter, and
Turb is turbidity, in formazin nephelometric units, measured with a YSI EXO2.

The calibration dataset generally spans the full range of observed turbidity values at the station, however there is limited model extrapolation. Extrapolation is defined as computation beyond the range of the model calibration dataset. Per USGS guidelines (USGS, 2016), the model may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model. The minimum and maximum turbidity values are shown below and the calibration data set covers 99% of the time-series record. The original maximum computed SSC was 407 mg/L (below) however is outside of the acceptable range of extrapolation. The portion of time-series data beyond the extrapolation limit is less than 1%. Following USGS guidelines, the extrapolated, maximum computed SSC for this model is limited to 131 mg/L.

Parameter	Minimum	Maximum
Turbidity (FNU) entire record	2.4	395
Computed SSC (mg/L)	6.1	*407/131

Suspended-Sediment Concentration Record

The complete SSC record is computed using this regression model and can be found at https://waterdata.usgs.gov/ca/nwis/uv/?site_no=11455146 as well as the links specifically to the stations in the sediment network at http://nrtwq.usgs.gov/ca.

Model

SSC = + 1.02 * TURB + 3.74

Variable Summary Statistics

	SSC	TURB
Minimum	8.0	3.31
1st Quartile	14.0	10.30
Median	20.0	17.80
Mean	34.5	30.10
3rd Quartile	51.0	44.90
Maximum	119.0	111.00

Basic Model Statistics

Number of Observations	27
Standard error (RMSE)	2.82
Average Model standard percentage error (MSPE)	8.16
Coefficient of determination (R^2)	0.992
Adjusted Coefficient of Determination (Adj. R^2)	0.991

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	3.74	0.7820	4.78	6.52e-05
TURB	1.02	0.0187	54.50	1.62e-27

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.721
E.vars	-0.721	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.222	0.193	0.544

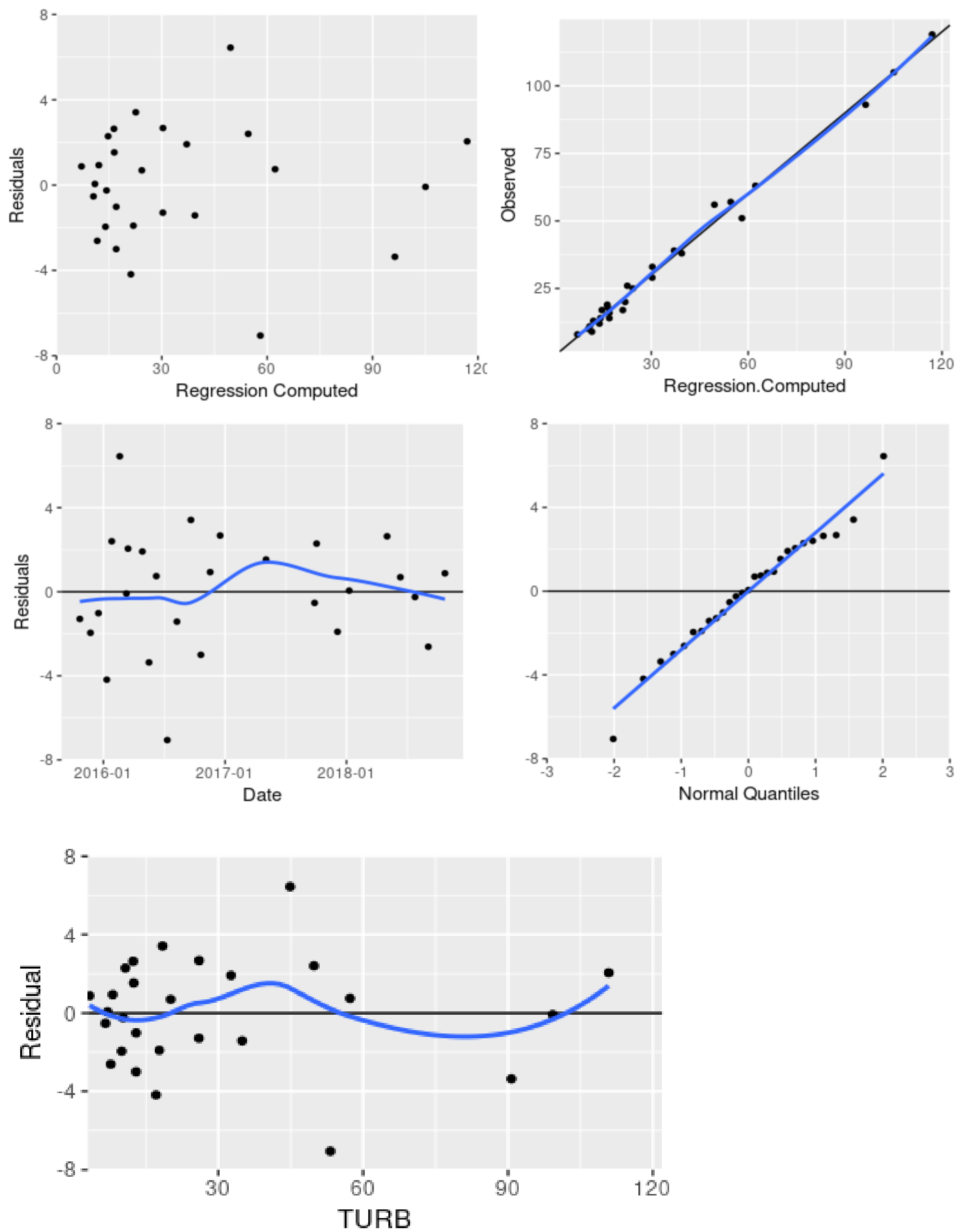
Flagged Observations

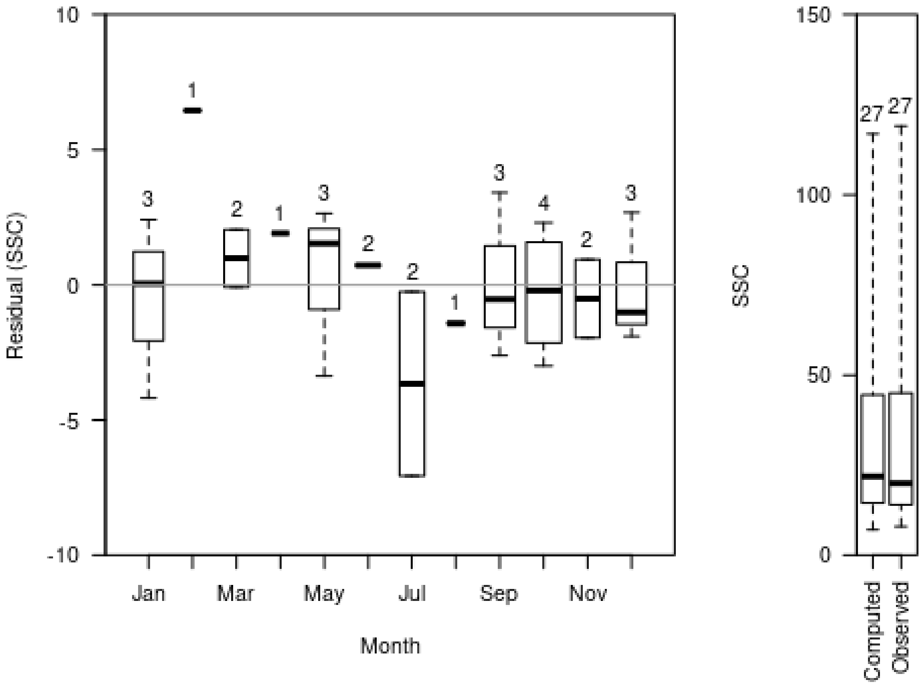
Date	Time	SSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's	DFFITS
2/19/2016	10:12	56	49.6	6.45	2.34	2.60	0.04	0.134	0.575
3/10/2016	11:44	105	105.0	-0.08	-0.03	-0.03	0.24	0.000	-0.019
3/15/2016	10:39	119	117.0	2.05	0.88	0.88	0.32	0.189	0.612
5/17/2016	10:27	93	96.4	-3.36	-1.33	-1.36	0.19	0.221	-0.676
7/11/2016	9:41	51	58.1	-7.06	-2.59	-2.96	0.06	0.215	-0.751

Plots:

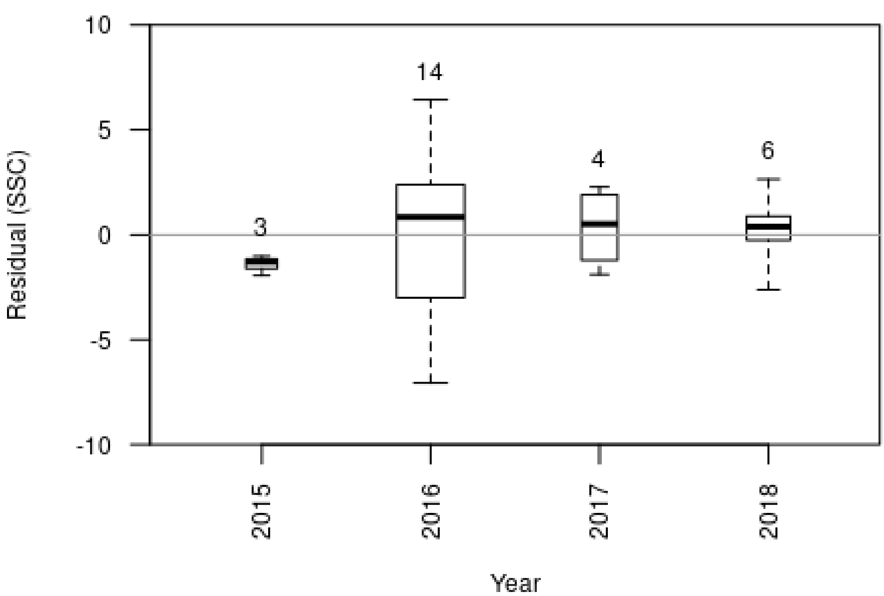
This summary is in accordance with the Office of Water Quality Technical Memorandum 2016.10 (USGS, 2016) which states this MAS **must** follow the format described in the memorandum. Based on this guidance, the following plots were generated using a specialized R-Script application specifically developed for this purpose by Patrick Eslick of the KSWSC (the MAS app) and is located at the following address:
https://patrickeslick.github.io/ModelArchiveSummary/

Statistical Plots and Residuals versus Time



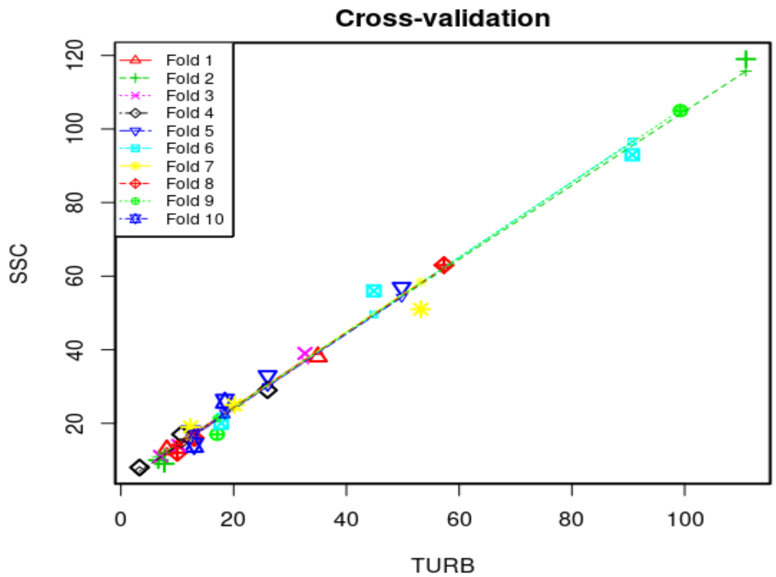


10
5
0
-5
-10
Residual (SSC)
Jan
Mar
May
Jul
Sep
Nov
Month
150
100
50
0
SSC
27
27
Computed
Observed

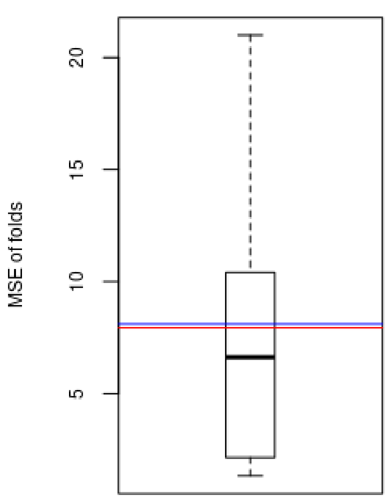


10
5
0
-5
-10
Residual (SSC)
3
14
4
6
2015
2016
2017
2018
Year

Cross-validation indicates when the model calibration data are randomly divided into subsets, the predictions from each subset regression model are very similar to the final regression model. The graph below shows a k-fold cross-validation with k=10 and the large points represent observations that were left out of each fold and are identified by the color and shape.




```
             Minimum MSE of folds:   1.33
                Mean MSE of folds:   8.11
              Median MSE of folds:   6.63
             Maximum MSE of folds:  21.00
 (Mean MSE of folds) / (Model MSE):   1.02
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Definitions

```
SSC: Suspended sediment concentration (SSC) in mg/l (80154)
TURB: Turbidity in FNU (63680)
```

Mas App Version 1.0

Model-Calibration Dataset

The calibration dataset table was compiled from values computed by SAID.

Observation Number	DateTime	SSC	TURB	Computed SSC	Residual	Normal Quantile	Censored Values
1	10/22/2015 10:22	29	26	30	-1.3	-0.478	--
2	11/23/2015 10:58	12	10	14	-2.0	-0.821	--
3	12/17/2015 10:26	16	13	17	-1.0	-0.377	--
4	1/11/2016 10:59	17	17.1	21	-4.2	-1.565	--
5	1/26/2016 13:44	57	49.8	55	2.4	0.958	--
6	2/19/2016 10:12	56	44.9	50	6.4	2.013	--
7	3/10/2016 11:44	105	99.2	105	-0.1	-0.092	--
8	3/15/2016 10:39	119	111	117	2.1	0.698	--
9	4/27/2016 10:00	39	32.7	37	1.9	0.585	--
10	5/17/2016 10:27	93	90.7	96	-3.4	-1.307	--
11	6/8/2016 8:25	63	57.3	62	0.7	0.185	--
12	7/11/2016 9:41	51	53.2	58	-7.1	-2.013	--
13	8/9/2016 10:18	38	34.9	39	-1.4	-0.585	--
14	9/20/2016 13:17	26	18.5	23	3.4	1.565	--
15	10/20/2016 11:36	14	13.0	17	-3.0	-1.115	--
16	11/17/2016 11:20	13	8.2	12	0.9	0.377	--
17	12/17/2016 10:58	33	26.0	30	2.7	1.307	--
18	5/4/2017 12:41	18	12.5	16	1.5	0.478	--
19	9/27/2017 13:23	10	6.6	11	-0.5	-0.280	--
20	10/3/2017 13:07	17	10.7	15	2.3	0.821	--
21	12/5/2017 13:53	20	17.8	22	-1.9	-0.698	--
22	1/9/2018 11:43	11	7.1	11	0.1	0.000	--
23	5/3/2018 11:09	19	12.3563	16	2.6	1.115	--
24	6/12/2018 12:22	25	20.1375	24	0.7	0.092	--
25	7/26/2018 8:48	14	10.29	14	-0.3	-0.185	--
26	9/4/2018 11:34	9	7.71	12	-2.6	-0.958	--
27	10/24/2018 9:27	8	3.31	7	0.9	0.280	--

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., http://dx.doi.org/10.3133/20151177.

Edwards, T.K. and Glysson, G.D., 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. C2, 89 p., Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. (Also available at http://pubs.usgs.gov/tm/3a23/.)

R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: https://www.R-project.org/.

Rasmussen, P.P., Gray J.R., Glysson G.D., Ziegler A.C., 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods book 3, chap. C4, 53 p., Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations. book 9, chap. A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

[USGS] U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.